# QUICKSORT WITH UNRELIABLE COMPARISONS: A PROBABILISTIC ANALYSIS

LAURENT ALONSO, PHILIPPE CHASSAING, FLORENT GILLET, SVANTE JANSON, EDWARD M. REINGOLD, AND RENÉ SCHOTT

Abstract. We provide a probabilistic analysis of the output of Quicksort when comparisons can err.

## 1. Introduction

Suppose that a sorting algorithm, knowingly or unknowingly, uses element comparisons that can err. Considering sorting algorithms based solely on binary comparisons of the elements to be sorted (algorithms such as insertion sort, selection sort, quicksort, and so on), what problems do we face when those comparisons are unreliable? For example, [6] gives a clever $\mathcal{O}\left(\epsilon^{-1}\log n\right)$ algorithm to assure, with probability $1-\epsilon$, that a putatively sorted sequence of length $n$ is truly sorted. But knowing the structure of the ill-sorted output would likely make error checking easier. Also, in situations in which a reliable comparison is the fruit of a long process, one could chose to interupt the comparison process, thus trading reliability of comparisons (and quality of the output) for time. As a first step in order to understand the consequences of errors, we propose to analyze the number of inversions in the output of a sorting algorithm (we choose Quicksort [10]) subject to errors.

We assume throughout this paper that the elements of the sequence

$$x = (x_1, x_2, \ldots, x_n)$$

to be sorted are distinct. We assume further that the only comparisons subject to error are those made between elements being sorted; that is, comparisons among indices and so on are always correct. Errors in element comparisons are random events, spontaneous and independent of each other, of position, and of value, with a common probability $p$, $n$ being the length of the list to be sorted. The number of inversions in the output sequence $y = (y_1, y_2, \ldots, y_n)$ is denoted

$$I(y) = \#\left\{(i,j) \mid 1 \le i < j \le n \text{ and } y_i > y_j\right\}.$$

We assume that the input list is presented in random order, each of the $n!$ random orders being equiprobable. Finally we denote by $I(n,p)$ the random number of inversions in the output sequence of Quicksort subject to errors.

Our result is, roughly speaking,

$$I(n,p) = \Theta\left(n^2 p\right),$$

when $(n,p) \to (\infty, c)$, meaning that $\frac{I(n,p)}{n^2 p}$ converges to some nondegenerate probability distribution. The "surprise", not so unexpected after the fact, is that there are phase changes in the limit law, depending on the asymptotic behaviour of $(n,p)$.

The organization of this paper is as follows: The results are stated in Section 2. In Section 3, we establish a general distributional identity for $I(n,p)$. In the remaining sections, we prove convergence results for $I(n,p)$ when:

- $p \to c$, $0 < c \le 1$,
- $p$ vanishes more slowly than $1/n$,
- $p \sim \lambda/n$ where $\lambda$ is a positive constant.

The case $np \to 0$ is different and not treated in detail; see Remark 2.10. In Section 4, we establish a general result of convergence using contraction methods (cf. [14, 15]), and we use it in Section 5, for the first two cases. These methods do not apply for Case 3, which requires poissonization (see Section 6, where we use an embedding of Quicksort in a Poisson point process).

## 2. Results

Set

$$X_{n,p} = \frac{I(n,p)}{n^2 p}.$$

We will always let $U$ denote a random variable that is uniformly distributed on $[0,1]$. Also, $\mathbb{N}^*$ shall denote the set of positive integers, and $\mathbb{N}$ the set of nonnegative integers.

**Case 1: $\lim p = c > 0$.**

**Theorem 2.1.** *If $\lim p = c$, $c \in (0,1]$, then $X_{n,p}$ converges in distribution to a random variable $X_c$ whose distribution is characterized as the unique solution with finite mean of the equation*

$$X_c \overset{law}{=} [(1-2c)U + c]^2 X_c + [(2c-1)U + 1 - c]^2 \widetilde{X}_c + T(c,U), \tag{1}$$

*in which $\widetilde{X}_c$ denotes a copy of $X_c$, $(X_c, \widetilde{X}_c, U)$ are independent, and*

$$T(c,U) = \frac{1-c}{2}(U^2 + (1-U)^2) + cU(1-U).$$

*Furthermore,*

$$\mathbb{E}\,[X_c] = \frac{2-c}{2(1+2c-2c^2)},$$

*and*

$$\mathrm{Var}\,(X_c) = \frac{(1-c)^2(1-2c)^2}{4(1+2c-2c^2)^2(3+6c-8c^2+4c^3-2c^4)}.$$

As usual with laws related to Quicksort, see e.g. [14, 15], $nU$ is approximately the position of the pivot of the first step of the algorithm. As in standard Quicksort recurrences, the coefficients of $X_c$ and of its independent copy $\tilde{X}_c$ are related to the sizes of the two sublists on the left and right of the pivot, sizes respectively asymptotic to $n\,((1-2c)U + c)$ and $n\,((2c-1)U + 1 - c)$. The toll function $T(c,U)$ is approximately $(n^2p)^{-1} \approx (n^2c)^{-1}$ times the number of inversions created in the first step: $c(1-c)n^2U^2/2$ is approximately the number of inversions of the $cnU$ elements, smaller than the pivot but misplaced on the right of it, with the $(1-c)\,nU$ elements smaller than the pivot, that are placed, as they should be, on the left; $c^2n^2U(1-U)$ is the number of inversions between misplaced elements from the two sides of the pivot. The toll function $T(c,U)$ depends on only one of the two

sources of randomness (the randomly ordered input list, and the places of the errors), viz., the first one, through $U$. The second source of randomness is killed by the law of large numbers: in the average, each of the $cnU+o(n)$ misplaced numbers from the right of the pivot produces inversions with one half of the $(1-c)nU+o(n)$ elements smaller than the pivot, that are placed, as they should be, on the left. As opposed to the other values of $c$, the choices $c=0.5$ and $c=1$ lead to deterministic $X_c = 1/2$, without any surprise : for $p = 0.5$ the output sequence is a random uniform permutation, with a number of inversions concentrated around $n^2/4$ [10, Chap. 5.1.1]; for $p = 1$ the output sequence is decreasing, and has $n(n-1)/2$ inversions.

**Case 2: $p$ vanishes more slowly than $\frac{1}{n}$.**

**Theorem 2.2.** *If $\lim p = 0$ and $\lim np = +\infty$, $X_{n,p}$ converges in distribution to a random variable $X$ whose distribution is characterized as the unique solution with finite mean of the equation*

$$X \overset{law}{=} U^2 X + (1-U)^2 \widetilde{X} + \frac{U^2+(1-U)^2}{2}. \tag{2}$$

*In (2), $\widetilde{X}$ denotes a copy of $X$ and $(X, \widetilde{X}, U)$ are independent. Furthermore,*

$$\mathbb{E}[X] = 1 \qquad \textit{and} \qquad \operatorname{Var}(X) = \frac{1}{12}.$$

Note that equation (2) is just (1) specialized to $c = 0$, but, as opposed to $c \neq 0$, an additional condition, $p \gg 1/n$, is needed to ensure that the law of large numbers still holds. Also, as another difference between (1) and (2), for $p \ll 1$ the errors do not change the sizes of the sublists in a significant way. The solution $X$ equals half the sum of the squares of the widths of the random intervals $[Y_{k,j}, Y_{k,j+1}]$ defined by (3) below. This is equivalent to the following statement:

**Proposition 2.3.** *The solution $X$ equals half the area $\int_0^1 Z(t)\,dt$ under the FIND limit process $Z$.*

For both these claims, see Remark 2.7. The Find process was introduced in [7] and is pictured at Figure 1.

**Case 3: $\lim np = \lambda$.**

Assume that

- $\Pi$ is a Poisson point process with intensity $\lambda$ on $\mathbb{N}^* \times [0,1]$, meaning that, for each $n$, $|\Pi \cap (\{n\} \times [0,1])|$ is a Poisson random variable with mean $\lambda$, and the second coordinates of points of $\Pi$ are uniform on $[0,1]$ and independent (see [9] for a general definition of Poisson point processes);
- $\{U_{k,j} : k \geq 0, 1 \leq j \leq 2^k\}$ is an array of independent uniform random variables on $[0,1]$, independent of $\Pi$;
- the random variables $(Y_{k,j}, k \geq 0, 1 \leq j \leq 2^k)$ are defined recursively by

$$\begin{aligned} Y_{0,0} = 0, \qquad Y_{0,1} = 1, \qquad Y_{k+1,2j} &= Y_{k,j} \qquad \text{for } 0 \leq j \leq 2^k, \\ Y_{k+1,2j-1} &= (1-U_{k,j})Y_{k,j-1} + U_{k,j}Y_{k,j} \qquad \text{for } 1 \leq j \leq 2^k; \end{aligned} \tag{3}$$

- for $x \in [0,1]$, $J_k(x) = 2j-1$ if $Y_{k-1,j-1} \leq x < Y_{k-1,j}$,

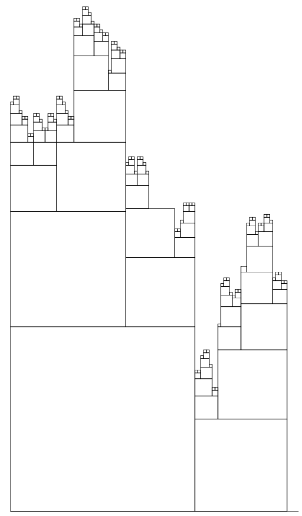

FIGURE 1. The Find process.

and define, for $\lambda > 0$, (the sum is a.s. finite by Lemma 6.4)

$$X(\lambda) = \frac{1}{\lambda} \sum_{(k,x)\in\Pi} \left| x - Y_{k,J_k(x)} \right|. \tag{4}$$

The variables $Y_{k,j}$ describe a fragmentation process (see [7] for historical references): we start with $[0,1)$ and recursively break each interval into two at a random point (uniformly chosen). In the $k$-th generation we thus have a partition of $[0,1)$ into $2^k$ intervals $I_{k,j}$, $1 \le j \le 2^k$, with $I_{k,j} = [Y_{k,j-1}, Y_{k,j})$. The interval of generation $k-1$ that contains $x$ is cut at step $k$ at the point $Y_{k,J_k(x)}$. Hence $\left|x - Y_{k,J_k(x)}\right|$ in (4) is the distance from $x$ to this cut point.

**Theorem 2.4.** *If* $\lim p = 0$ *and* $\lim np = \lambda > 0$, *then* $X_{n,p}$ *converges in distribution to* $X(\lambda)$. *The family* $\{X(\lambda)\}_{\lambda>0}$ *of random variables satisfies the distributional identity:*

$$X(\lambda) \overset{law}{=} U^2 X(\lambda U) + (1-U)^2 \widetilde{X}(\lambda(1-U)) + \Theta(\lambda, U), \tag{5}$$

*in which, conditionally given that* $U = u$, $X(\lambda U)$, $\widetilde{X}(\lambda(1-U))$ *and* $\Theta(\lambda, U)$ *are independent,* $X(\lambda U)$ *and* $\widetilde{X}(\lambda(1-U))$ *are distributed as* $X(\lambda u)$ *and* $X(\lambda(1-u))$, *respectively, and*

$$\Theta(\lambda, u) \overset{law}{=} \frac{1}{\lambda} \sum_{i=1}^{N_\lambda} |u - V_i|,$$

*in which* $N_\lambda$ *is a Poisson random variable with mean* $\lambda$, *the random variables* $V_i$ *are uniformly distributed on* $[0,1]$, *and* $N_\lambda$, *and the* $V_i$*'s are independent. Furthermore,*

$$\mathbb{E}[X(\lambda)] = 1, \qquad \operatorname{Var}(X(\lambda)) = \frac{1}{12} + \frac{1}{3\lambda}. \tag{6}$$

**Remark 2.5.** The distributional identities (1), (2) and (5) really are equations for distributions, but it is more convenient to state them for random variables

as done here. For (5) to make sense, i.e. in order to insure, for instance, that $X(\lambda U)$ is a random variable, it is implicitly assumed that the random variables $X(\lambda)$ depend measurably on $\lambda$. Thus a solution of (5) is a family of probability measures $\mu = (\mu_\lambda)_{\lambda>0}$ on $[0,+\infty)$, such that there exists a family $Y = (Y(\lambda))_{\lambda>0}$ of random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and satisfying the following properties:

(i) for $\lambda > 0$, $\mu_\lambda$ is the distribution of $Y(\lambda)$,

(ii) $Y$ is a measurable process [8, Chap. 1], meaning that the mapping

$$(\lambda, \omega) \to Y(\lambda, \omega) : \quad ((0,+\infty) \times \Omega, \mathcal{B}((0,+\infty)) \otimes \mathcal{A}) \to ([0,+\infty), \mathcal{B}([0,+\infty)))$$

is measurable,

and such that (5) holds for $Y$. A measurable version of the stochastic process $X = (X(\lambda))_{\lambda>0}$ is defined at (8) below (measurability follows from [8, Rem. 1.14]).

For uniqueness, we need extra assumptions: let $\mathcal{M}$ denote the class of families of distributions $\mu = (\mu_\lambda)_{\lambda>0}$ satisfying (i) and (ii) above, plus the condition:

(iii) for some $\alpha \in (0,1)$, the function

$$\lambda \longrightarrow \lambda^\alpha \mathbb{E}\left[Y(\lambda)\right]$$

is bounded on any bounded interval of $(0,+\infty)$.

Let $\nu_\lambda$ denote the distribution of $X(\lambda)$. We have

**Theorem 2.6.** *The family $\nu = (\nu_\lambda)_{\lambda>0}$ is the unique solution of* (5) *in $\mathcal{M}$.*

We do not know whether the extra assumption (iii) is necessary. Let us comment further on equation (4). Writing $\Pi_k = \{x : (k,x) \in \Pi\}$ and $\Pi_{k,j} = \Pi_k \cap I_{k,j}$, we can thus rewrite (4) as

$$X(\lambda) = \frac{1}{\lambda} \sum_{k=1}^{\infty} \sum_{j=1}^{2^k} \sum_{x \in \Pi_{k,j}} |x - x_{k,j}|, \tag{7}$$

where $x_{k,j}$ is either the left or right endpoint of $I_{k,j}$ (depending on whether $j$ is even or odd).

Note that, conditioned on the partitions $\{I_{k,j}\}$, i.e. on $\{Y_{k,j}\}_{k,j}$, each $\Pi_{k,j}$ is a Poisson process on $I_{k,j}$ with intensity $\lambda$, with the processes $\Pi_{k,j}$ independent. Since only the distribution of $X(\lambda)$ matters, we can by this conditioning and an obvious symmetry of the Poisson processes $\Pi_{k,j}$ just as well let $x_{k,j}$ in (7) be the left endpoint of $I_{k,j}$ for every $k$ and $j$.

Let $\Pi'$ be a Poisson process on $(0,1] \times (0,\infty)$ with intensity 1, and let $\xi(t) = \sum_{(x,y)\in\Pi',\, y\le t} x$, $t \ge 0$. (This is a pure jump Lévy process with Lévy measure $\mathbb{1}_{(0,1]}dt$.) Let $\xi^{(k,j)}(t)$ be independent copies of this process, independent of $\{Y_{k,j}\}$. A scaling argument shows that (7) can be written

$$X(\lambda) = \frac{1}{\lambda} \sum_{k=1}^{\infty} \sum_{j=1}^{2^k} |I_{k,j}| \xi^{(k,j)}(\lambda |I_{k,j}|). \tag{8}$$

**Remark 2.7.** Let $\hat{X} = \frac{1}{2} \sum_{k=1}^{\infty} \sum_{j=1}^{2^k} |I_{k,j}|^2$. Then $\hat{X}$ satisfies (2), so that $\hat{X}$ is the limit variable $X$ in Theorem 2.2. ($X$ is a.s. finite and has finite mean by Lemma 6.1.) Moreover, the FIND limit process $Z$ in [7] is defined by $Z(t) =$

$\sum_{k=1}^{\infty}\sum_{j=1}^{2^k}|I_{k,j}|\mathbb{1}_{t\in I_{k,j}}$; hence $\int_0^1 Z(t)\,dt = \sum_{k=1}^{\infty}\sum_{j=1}^{2^k}|I_{k,j}|^2 = 2X$. This justifies Proposition 2.3.

Moreover, by the law of large numbers, $\mathbb{E}|\lambda^{-1}\xi(\lambda) - 1/2| \to 0$ as $\lambda \to \infty$. It follows (by dominated convergence using Lemma 6.1) that for the special version of $X(\lambda)$ defined at (8)

$$\mathbb{E}|X(\lambda) - X| \to 0,$$

and hence $X(\lambda)$ converges to $X$ in distribution as $\lambda \to \infty$.

In this third case, we have a system of equations involving an infinite family of laws, and we could not adapt the contraction method: we rather use a poissonization. The phase transition from (2) to (5) is explained easily: instead of a number of errors $\gg 1$, we have now $\mathcal{O}(1)$ errors at each step, and the law of large numbers does not hold anymore for the number of inversions produced by step 1. Actually the number $N_\lambda$ of errors at the first step is asymptotically Poisson distributed, and the $N_\lambda$ errors are at positions $nV_i$, approximately uniformly distributed on $[0,1]$. Thus, the number of inversions caused by this first step is approximately

$$n\sum_{i=1}^{N_\lambda}|U - V_i| \approx n^2 p\Theta(\lambda, U).$$

**Remark 2.8.** Actually we prove a stronger theorem in each of the three cases, as we prove convergence of laws for the Wasserstein $d_1$ metric [13]. It entails convergence of the first moment. The convergence of higher moments is an open problem.

**Remark 2.9.** As we shall see in Section 6, the distribution tail $\mathbb{P}(X(\lambda) \geq x)$ decreases exponentially fast (Theorem 6.5).

**Remark 2.10.** When $np \to 0$ very slowly, that is $(np)^{-1} \ll \log n$, we conjecture that $2np\log(I(n,p)/n)$ converges in distribution to $\log U$, with the consequence that $n^{1-\varepsilon} \ll I(n,p) \ll n$, for any positive $\varepsilon$. Actually, the main contribution to $I(n,p)$ comes from the "first" error, in some sense. When $(np)^{-1} \sim \log n$, the probability that no error occurs has a positive limit: we conjecture that, conditionally given the occurence of at least one error, the situation is similar to the previous case, that is, $\log(I(n,p))\,/\log n$ converges in distribution to a random variable with values in $(0,1)$. When $(np)^{-1} \gg \log n$, $\mathbb{P}(I(n,p) = 0) \to 1$.

**Remark 2.11.** Finally, we would like to stress that in the proof of convergence for one the three regimes considered in this Section, we have to deal simultaneously with any sequence $(n, p_n)$ converging to $(+\infty, c)$ according to this regime. This can be observed on the key equation (9), for instance, in which we would like to argue, roughly speaking, that if $(n,p)$ is close to $(+\infty, c)$ according to a given regime, then $(Z_{n,p} - 1, p)$ and $(n - Z_{n,p}, p)$ are also close to $(+\infty, c)$ according to the same regime, with a large probability: here the same probability $p$ is associated to three different integers, $n$, $Z_{n,p} - 1$ and $n - Z_{n,p}$, that denote the sizes of the input list, and of the two sublists formed at the first step of Quicksort, respectively. Thus $p$ cannot be seen as a sequence indexed by $n$. In order to allow such a loose relation between $n$ and $p$, filters turn out to be more handy than sequences (see [2, Chap. I]). Convergences in the three regimes are thus understood as convergences along the three corresponding filters (see Theorem 4.2).

## 3. A distributional identity for the number of inversions

At the first step Quicksort compares all elements of the input list with the first element of the list (usually called *pivot*). All items less (resp. larger) than the pivot are stored in a sublist on the left (resp. right) of the pivot. Comparisons are not reliable, therefore $s_\ell$ items that should belong to the left sublist are wrongly stored in the right sublist, and $s_r$ items larger than the pivot are misplaced in the left sublist.

Since its items are chosen randomly, the input list is a random permutation and the true rank of the pivot can be written $\lceil nU \rceil$, where $U$ is uniformly distributed on $[0, 1]$ and $\lceil x \rceil$ is the ceiling of $x$. Also, conditionally given $U$, $s_\ell$ (resp. $s_r$) is a binomial random variable with parameters $(\lceil nU \rceil - 1,\, p)$ (resp. $(n - \lceil nU \rceil,\, p)$). Quicksort with error is then independently applied on the left sublist $\ell$ and on the right sublist $r$ and new errors occur, ultimately producing two new sublists $\tilde{\ell}$ and $\tilde{r}$. Set

$$Z_{n,p} = \lceil nU \rceil - s_\ell + s_r,$$

so that $Z_{n,p} - 1$ (resp. $n - Z_{n,p}$) is the size of $\ell$ and $\tilde{\ell}$ (resp. $r$ and $\tilde{r}$).

In order to enumerate the inversions of the output list, we introduce a *purely fictitious* error-correcting algorithm that parallels the implementation of Quicksort: This fictitious error-correcting algorithm has two recursive steps,

- First, the error-correcting algorithm corrects the sublists $\tilde{\ell}$ (resp. $\tilde{r}$) at costs $L = I(\tilde{\ell})$ (resp. $R = I(\tilde{r})$), producing two increasing sublists $\hat{\ell}$ and $\hat{r}$. Note that $L$ and $R$ are conditionally independent, given $Z_{n,p}$. Furthermore, the two sublists $\ell$ and $r$ obtained at the end of Step 1 are in uniform random order before the second step of Quicksort, so that, conditionally given $Z_{n,p}$, cost $L$ (resp. $R$) is distributed as $I(Z_{n,p} - 1, p)$ (resp. $I(n - Z_{n,p}, p)$).
- Then the error-correcting algorithm corrects the errors of Step 1, at a cost $t(n, p) = I(\hat{\ell} \| \text{pivot} \| \hat{r})$. Here $\hat{\ell} \| \text{pivot} \| \hat{r}$ stands for the list obtained when one puts $\hat{\ell}$, the pivot and $\hat{r}$ side by side. The number of inversions $t(n, p)$ in the list $\hat{\ell} \| \text{pivot} \| \hat{r}$ is analyzed in detail at the end of this section.

These two steps lead to the following equation for $I(n, p)$:

$$I(n, p) \overset{law}{=} I(Z_{n,p} - 1, p) + I'(n - Z_{n,p}, p) + t(n, p) \tag{9}$$

where $Z_{n,p} = \lceil nU \rceil - s_\ell + s_r$. We shall obtain the asymptotic distribution of $t(n, p)$, and as a consequence (9) will translate, after renormalisation, into a distributional identity satisfied by the limit law of $I(n, p)/(n^2 p)$. The limit law appears on both sides of the distributional identity, as expected, due to the recursive structure of Quicksort, and is thus characterized as the fixed point of some transformation.

**Description of $t(n, p)$.** At the end of the first step of the error-correcting algorithm, we obtain two subarrays $\hat{\ell}$ and $\hat{r}$, left and right of the pivot (cf. Figure 3). They are sorted in increasing order but there are $s_r$ (red) elements larger than the pivot just to its left and $s_\ell$ (green) elements smaller than the pivot element just to its right. Thus, the only misplaced elements that the proofreader must correct in step 2 are clustered around the pivot.

In order to sort the list, the red and green sublists must be exchanged. This requires $s_\ell s_r + s_\ell + s_r$ inversions. We get therefore two unsorted lists $\vec{\ell}$ and $\vec{r}$ each composed of two sorted sublists. All items of $\vec{\ell}$ (resp. of $\vec{r}$ ) are now smaller

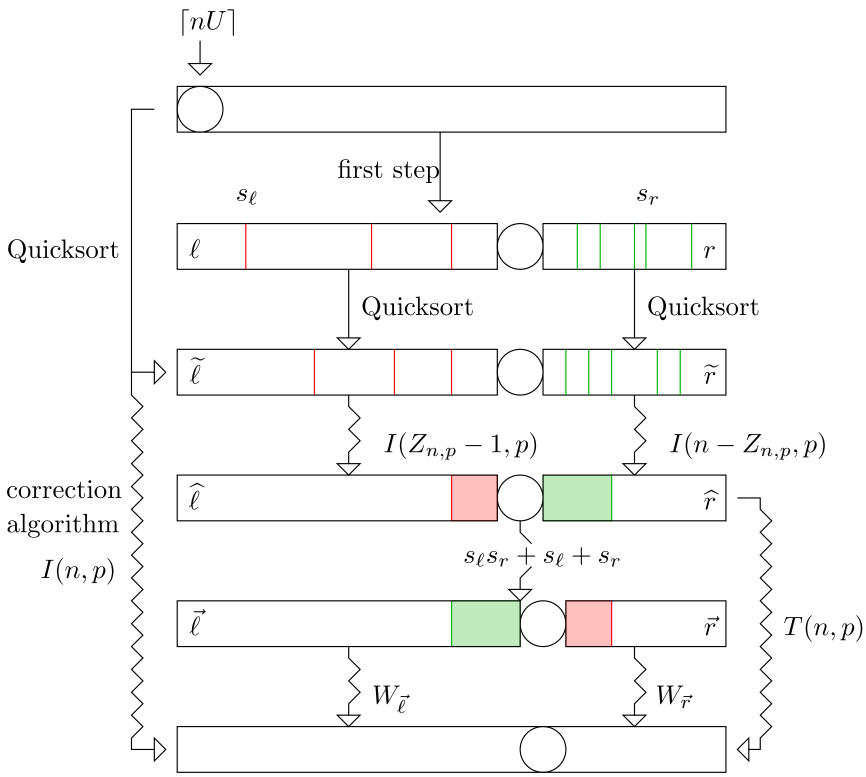


FIGURE 2. The error-correcting algorithm.

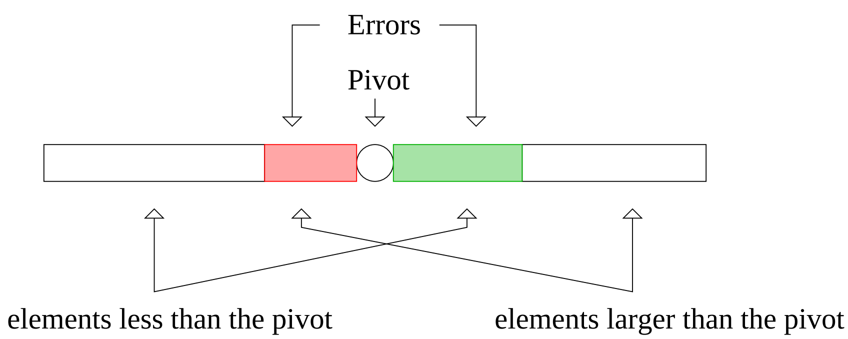


FIGURE 3. The two sublists $\hat{\ell}$ and $\hat{r}$.

(resp. larger) than the pivot, so that the length of $\vec{\ell}$ (resp. of $\vec{r}$ ) is $\lceil nU \rceil - 1$ (resp. $n - \lceil nU \rceil$). It remains to sort $\vec{\ell}$ and $\vec{r}$, at respective costs $W_{\vec{\ell}}$ and $W_{\vec{r}}$ that are conditionally independent given $U$, leading to:

$$t(n,p) = s_\ell s_r + s_\ell + s_r + W_{\vec{\ell}} + W_{\vec{r}}. \tag{10}$$

**A model for $\left(W_{\vec{\ell}}, W_{\vec{r}}\right)$.** Let $W_m$ be the number of inversions in a list of $m$ elements sorted as follows: each element is painted black (white) with probability $p$ (resp. $1-p$). Then the black and white sublists are separately sorted in increasing order and the two sorted sublists are placed side by side, producing a new list $h$ with $m$ elements. We have

**Proposition 3.1.** *Let $Y_1, \ldots, Y_m$ be $m$ independent Bernoulli random variables with the same parameter $p$, and let $S_m = Y_1 + \ldots + Y_m$. Then*

$$W_m \overset{law}{=} \Big(\sum_{i=1}^{m} iY_i\Big) - \frac{S_m(S_m+1)}{2}.$$

*Proof.* Let us abbreviate $S_m$ to $S$. Among the $Y_i$'s, let $Y_{i_1}, \ldots, Y_{i_S}$ denote the $S$ random variables equal to 1, $Y_{i_{S+1}}, \ldots, Y_{i_m}$ those equal to 0, with $i_1 < \ldots < i_S$ and $i_{S+1} < \ldots < i_m$. Now $W_m$ can be seen as the number of inversions of the list $\left(i_j\right)_{1\le j\le m}$. In order to move the numbers $i_j$ ($j \le S$) to the correct position, the proofreader corrects inversions with each of the $i_j - j$ elements of $\{1, \ldots, m\}$ that are smaller than $i_j$ and do not belong to $\{i_1, \ldots, i_S\}$. Thus

$$W_m = \sum_{j=1}^{S} (i_j - j), \tag{11}$$

leading to the result. □

With the help of Proposition 3.1, we can give a useful description of the distribution of $(s_\ell, W_{\vec{\ell}})$ and $(s_r, W_{\vec{r}})$:

**Proposition 3.2.** *Conditionally given that the length of $\vec{\ell}$ is $m-1$, $(s_\ell, W_{\vec{\ell}})$ and $(s_r, W_{\vec{r}})$ are independent and distributed as $(S_{m-1}, W_{m-1})$ and $(S_{n-m}, W_{n-m})$, respectively.*

To sum up the results of this section, renormalizing (9), one obtains a distributional identity satisfied by $X_{n,p}$:

$$X_{n,p} \overset{law}{=} A_{n,p} X_{Z_{n,p}-1,p} + B_{n,p} \widetilde{X}_{n-Z_{n,p},p} + T_{n,p} \tag{12}$$

in which

$$Z_{n,p} = \lceil nU \rceil - s_\ell + s_r, \tag{13}$$

$$A_{n,p} = \left(\frac{Z_{n,p}-1}{n}\right)^2, \tag{14}$$

$$B_{n,p} = \left(\frac{n - Z_{n,p}}{n}\right)^2, \tag{15}$$

$$t(n,p) = s_\ell s_r + s_\ell + s_r + W_{\vec{\ell}} + W_{\vec{r}}, \tag{16}$$

$$T_{n,p} = \frac{t(n,p)}{n^2 p}, \tag{17}$$

and

- $U$ is a uniform random variable on $[0,1]$, and $\lceil nU \rceil$ is the position of the pivot,
- conditionally given $\lceil nU \rceil = m$, $(s_\ell, W_{\vec{\ell}})$ and $(s_r, W_{\vec{r}})$ are distributed as in Proposition 3.2,

- $X = (X_m)_{m\geq 0}$, $\widetilde{X} = \left(\widetilde{X}_m\right)_{m\geq 0}$ are two independent sequences with the same (unknown) distribution, independent of $(U, s_\ell, W_{\vec{\ell}}, s_r, W_{\vec{r}})$, and therefore of $(A_{n,p}, B_{n,p}, Z_{n,p}, T_{n,p})$.

The errors having a balancing effect: $Z_{n,p} = \lceil nU \rceil - s_\ell + s_r$ has the same mean, $(n+1)/2$, and a smaller variance than $\lceil nU \rceil$. We prove this in the following form.

**Lemma 3.3.**

$$\mathbb{E}\left[(Z_{n,p}-1)^2 + (n-Z_{n,p})^2\right] \leq \mathbb{E}\left[(\lceil nU \rceil - 1)^2 + (n - \lceil nU \rceil)^2\right] = \frac{(n-1)(2n-1)}{3} \leq \frac{2}{3}n^2.$$

*Proof.* The left hand side is the expected number of ordered pairs $(i,j)$ that end up on a common side of the pivot. This happens if $i$ and $j$ originally are on the same side of the pivot and we either compare both correctly or make errors for both of them, or if they are on opposite sides of the pivot and we make an error for exactly one of them. Hence

$$\begin{aligned}\mathbb{E}\left[(Z_{n,p}-1)^2 + (n-Z_{n,p})^2\right] &= \big(p^2 + (1-p)^2\big)\mathbb{E}\left[(\lceil nU \rceil - 1)^2 + (n - \lceil nU \rceil)^2\right] \\ &\quad + 2p(1-p)2\mathbb{E}\left[(\lceil nU \rceil - 1)(n - \lceil nU \rceil)\right] \\ &= \mathbb{E}\left[(\lceil nU \rceil - 1)^2 + (n - \lceil nU \rceil)^2\right] - 2p(1-p)\mathbb{E}\left[(\lceil nU \rceil - 1 - (n - \lceil nU \rceil))^2\right]\end{aligned}$$

which proves the first inequality. The rest is a simple calculation. □

Let us say that an element $a$ of the list, or the comparison in which $a$ plays the rôle of pivot, has depth $k$ if $a$ experiences $k-1$ comparisons before playing the rôle of pivot. We assume in this Section that any comparison with depth $k+1$ is performed after the last comparison with depth $k$. We call step $k$ the set of comparisons with depth $k$, and we let $I^{(k)}(n,p)$ denote the number of inversions created at step $k$, that is, the total number of inversions, in the output, between elements that are still in the same sublist before step $k$, but are not in the same sublist after step $k$. We shall need the following bound:

**Lemma 3.4.** *For every $k \geq 1$,*

$$\mathbb{E}\left[I^{(k)}(n,p)\right] \leq \frac{1}{2}\left(\frac{2}{3}\right)^k n^2 p.$$

*Proof.* For $k=1$, $I^{(1)}(n,p) = t(n,p)$, and a simple calculation yields

$$\mathbb{E}\left[t(n,p)\right] = p\frac{(n-1)(n+1)}{3} - p^2\frac{(n-1)(n-2)}{6} \leq \frac{1}{3}n^2 p.$$

For $k>1$ we find by induction, conditioning on the partition in the first step,

$$\mathbb{E}\left[I^{(k)}(n,p)\right] \leq \mathbb{E}\left[\frac{1}{2}\left(\frac{2}{3}\right)^{k-1}(Z_{n,p}-1)^2 p + \frac{1}{2}\left(\frac{2}{3}\right)^{k-1}(n-Z_{n,p})^2 p\right]$$

and the result follows by Lemma 3.3. □

**Proposition 3.5.** *Set $a_{n,p} = \mathbb{E}\left[X_{n,p}\right]$. Then*

$$a_{n,p} \leq 1.$$

*Proof.* By Lemma 3.4, $a_{n,p} \leq \sum_1^\infty \frac{1}{2}\left(\frac{2}{3}\right)^k$. □

## 4. Fixed point theorems

The proofs of the first two cases are examples of the contraction method [14, 15]: on one hand we have more or less explicitly defined random variables $A^{(i)}_{n,p}$, $1 \le i \le I$, and $T_{n,p}$, and we know how to prove directly that they converge to $A^{(i)}$, $T$. On the other hand, we have a family $X_{n,p}$ of random variables defined by induction:

$$X_{n,p} \overset{\text{law}}{=} \sum_{i=1}^{I} A^{(i)}_{n,p} X^{(i)}_{Z^{(i)}_{n,p},p} + T_{n,p}, \tag{18}$$

and a random variable $X$ implicitly defined by the distributional identity

$$X \overset{\text{law}}{=} \sum_{i=1}^{I} A^{(i)} X^{(i)} + T, \tag{19}$$

in which, in some sense, $\lim Z^{(i)}_{n,p} = +\infty$. Then, under additional technical conditions, the convergence of the "coefficients" $A^{(i)}_{n,p}$, $T_{n,p}$, entails the convergence of the "solution" $X_{n,p}$. One has to prove existence and unicity of the solutions, usually as fixed points of contracting transformations in a subspace of the space of probability measures, with a suitable metric. In the case we are interested in, (18) holds and:

- $I$ is a fixed positive integer;
- $C_{n,p} = (A^{(1)}_{n,p}, Z^{(1)}_{n,p}, \dots, A^{(I)}_{n,p}, Z^{(I)}_{n,p}, T_{n,p})$ is a given random vector for each $n, p$;
- $Z^{(i)}_{n,p} \in [0, \dots, n-1]$;
- The families $(X^{(i)}_{n,p})_{n,p}$, $i = 1, 2, \dots, I$, are i.i.d. and independent of $C_{n,p}$, and and $X^{(i)}_{n,p} \overset{law}{=} X_{n,p}$.

Given such $C_{n,p}$ we thus define, for any distributions $G_{0,p}, \dots, G_{n-1,p}$,

$$\Phi(G_{0,p}, \dots, G_{n-1,p}) = \mathcal{L}\left( \sum_{i=1}^{I} A^{(i)}_{n,p} X^{(i)}_{Z^{(i)}_{n,p},p} + T_{n,p} \right),$$

when, as above, the families $(X^{(i)}_{k,p})_{k,p}$, $i = 1, 2, \dots, I$, are i.i.d. and independent of $C_{n,p}$, and further $X^{(i)}_{k,p}$ has the distribution $G_{k,p}$. Thus (18) can be written

$$G_{n,p} = \Phi(G_{0,p}, \dots, G_{n-1,p}).$$

For (19) we similarly assume

- $C = (A^{(1)}, \dots, A^{(I)}, T)$ is a given random vector;
- the variables $X^{(i)}$, $i = 1, 2, \dots, I$ are i.i.d. and independent of $C$, and $X^{(i)} \overset{law}{=} X$.

Given such $C$ we define

$$\Psi(F) = \mathcal{L}\left( \sum_{i=1}^{I} A^{(i)} X^{(i)} + T \right),$$

when the variables $X^{(i)}$, $i = 1, 2, \dots, I$ are i.i.d. with distribution $F$ and independent of $C$. Then (19) can be written

$$\Psi(F) = F.$$

Let $D$ be the space of probability measures $\mu$ on $\mathbb{R}$ such that $\int_{\mathbb{R}} |x| \; d\mu(x) < +\infty$. The space $D$ is endowed with the Wasserstein metric

$$d_1(\mu,\nu) = \inf_{\substack{\mathcal{L}(X)=\mu \\ \mathcal{L}(Y)=\nu}} \|X-Y\|_1 = \left\|F^{-1}(U) - G^{-1}(U)\right\|_1 . \tag{20}$$

in which $F$ and $G$ denote the distribution functions of $\mu$ and $\nu$, $F^{-1}$ (resp. $G^{-1}$) denote the generalized inverses of $F$ and $G$ and, as in previous sections, $U$ is a uniform random variable [4]. Since $F^{-1}(U)$ (resp. $G^{-1}(U)$) has distribution $\mu$ (resp. $\nu$), the infimum is attained in relation (20).

The metric $d_1$ makes $D$ a complete metric space. Convergence of $\mathcal{L}(X_n)$ to $\mathcal{L}(X)$ in $D$ is equivalent to convergence of $X_n$ to $X$ in distribution *and*

$$\lim \mathbb{E}\left[|X_n|\right] = \mathbb{E}\left[|X|\right].$$

Therefore convergence in $D$ entails

$$\lim \mathbb{E}\left[X_n\right] = \mathbb{E}\left[X\right].$$

We refer to [13] for an extensive treatment of Wasserstein metrics. In what follows, we shall improperly refer to the convergence of $X_n$ to $X$ in $D$, meaning the convergence of their distributions. Let us take care first of relation (19):

**Theorem 4.1.** *If* $\sum_{i=1}^{I} \mathbb{E}\left[\left|A^{(i)}\right|\right] < 1$ *and* $\mathbb{E}\left[|T|\right] < \infty$*, then* $\Psi$ *is a strict contraction and (19) has a unique solution in* $D$.

*Proof.* Let $(X,Y)$ be a coupling of random variables, with laws $\mu$ and $\nu$, respectively, such that

$$\mathbb{E}\left[|X-Y|\right] = d_1(\mu,\nu).$$

Let $\left(\left(X^{(i)},Y^{(i)}\right)\right)_{1\le i\le I}$ be $I$ independent copies of $(X,Y)$. Furthermore, assume that $C$ and $\left(\left(X^{(i)},Y^{(i)}\right)\right)_{1\le i\le I}$ are independent. Then the probability distribution of

$$\sum_{i=1}^{I} A^{(i)}X^{(i)} + T, \quad \text{resp. } \sum_{i=1}^{I} A^{(i)}Y^{(i)} + T$$

is $\Psi(\mu)$ (resp. $\Psi(\nu)$) and

$$\begin{aligned} d_1(\Psi(\mu),\Psi(\nu)) &\le \sum_{i=1}^{I} \mathbb{E}\left[\left|A^{(i)}\right| \left|X^{(i)} - Y^{(i)}\right|\right] \\ &\le d_1(\mu,\nu) \sum_{i=1}^{I} \mathbb{E}\left[\left|A^{(i)}\right|\right]. \end{aligned}$$

Thus $\Psi$ is a contraction with contraction constant smaller than 1. Since $D$ is a complete metric space, this implies that $\Psi$ has a unique fixed point in $D$, by Banach's fixed point theorem. $\square$

We prove now a theorem which is a variant of those used by the previously cited authors: the difference is not deep, but here we deal with family of laws, not sequences, as we have two parameters, $n$ and $p$. As a consequence, to cover

Theorems 2.1 and 2.2, it will be convenient in their proofs to consider convergence with respect to a *filter* $\mathcal{F}$ on $\mathbb{N} \times [0,1]$, see [2, Chap. 1]. The collection of sets

$$V_{N,\varepsilon} = \{n \geq N\} \times ([c-\varepsilon, c+\varepsilon] \cap [0,1]), \qquad N \geq 0,\ \varepsilon > 0,$$

is a basis for the filter $\mathcal{F}_1$ corresponding to Theorem 2.1, while

$$\widetilde{V}_{N,\varepsilon} = \left\{(n,p) \,\middle|\, 0 < p \leq \varepsilon,\, n \geq N/p\right\}, \qquad N \geq 0,\ \varepsilon > 0,$$

is a basis for the filter $\mathcal{F}_2$ corresponding to Theorem 2.2.

**Theorem 4.2.** *Suppose that* (18) *holds for* $n \geq 1$ *and* $X_{0,p} = 0$*; i.e.* $G_{n,p} = \Phi(G_{0,p}, \ldots, G_{n-1,p})$ *for* $n \geq 1$ *and* $G_{0,p} = \delta_0$*, where* $G_{n,p} = \mathcal{L}(X_{n,p})$*. If*

i) $\left(\mathbb{E}\left[X_{n,p}\right]\right)_{n,p}$ *is bounded,*

ii) $\displaystyle\sum_{i=1}^{I} \mathbb{E}\left[\left|A^{(i)}\right|\right] < 1,$

iii) $T_{n,p} \xrightarrow[\mathcal{F}]{L^1} T,\ A^{(i)}_{n,p} \xrightarrow[\mathcal{F}]{L^1} A^{(i)},$

iv) $\lim_{\mathcal{F}}\ \mathbb{E}\left[\left|A^{(i)}_{n,p}\right|\ ;\ (Z^{(i)}_{n,p}, p) \notin V\right] = 0, \qquad \forall V \in \mathcal{F},$

*then* $X_{n,p}$ *converges in distribution to* $F$*, the unique solution of the equation* $\Psi(F) = F$ *in* $D$*. More precisely,* $d_1(G_{n,p}, F) \to 0$ *along* $\mathcal{F}$*.*

We need a lemma before proving Theorem 4.2.

**Lemma 4.3.** *Assume that three families of nonnegative numbers* $(a_{n,p})_{0 \leq n, 0<p<1}$*,* $(b_{n,p})_{0 \leq n, 0<p<1}$*, and* $(\gamma_{i,n,p},\ 0 \leq n, 0 \leq i \leq n, 0 < p < 1)$ *satisfy the inequalities:*

$$a_{n,p} \leq b_{n,p} + \sum_{i=0}^{n-1} \gamma_{i,n,p} a_{i,p},$$

*Let* $\mathcal{F}$ *be a filter. Under the following assumptions:*

- $a_{n,p}$ *is nonnegative and bounded,*
- *for some* $\Gamma < 1$ *and some* $V_0 \in \mathcal{F}$*,* $\forall (n,p) \in V_0,\ \sum_{k=0}^{n-1} \gamma_{k,n,p} < \Gamma$*,*
- $\lim_{\mathcal{F}} b_{n,p} = 0$*,*
- $\forall V \in \mathcal{F},\ \lim_{\mathcal{F}} \sum_{k:(k,p)\notin V} \gamma_{k,n,p} = 0$*,*

*we have*

$$\lim_{\mathcal{F}}\ a_{n,p} = 0.$$

*Proof of Lemma 4.3.* The proof is a variant of the proof of [14, Proposition 3.3]. Let $M$ be a bound for $a_{n,p}$, and let

$$a = \limsup_{\mathcal{F}}\ a_{n,p}.$$

For any $\epsilon > 0$, let $V_\epsilon \in \mathcal{F}$ be such that for $(n,p) \in V_\epsilon$,

$$a_{n,p} \leq a + \epsilon.$$

Then for $(n,p) \in V_\epsilon \cap V_0$ we have

$$\begin{aligned} a_{n,p} \quad &\leq \quad \sum_{k:(k,p)\notin V_\epsilon} \gamma_{k,n,p} a_{k,p} + \sum_{k:(k,p)\in V_\epsilon} \gamma_{k,n,p} a_{k,p} + b_{n,p} \\ &\leq \quad M \sum_{k:(k,p)\notin V_\epsilon} \gamma_{k,n,p} + (a+\epsilon)\Gamma + b_{n,p}. \end{aligned}$$

Taking lim sups, we obtain that for any $\epsilon > 0$,

$$a \leq (a+\epsilon)\Gamma.$$

Thus $a \leq a\Gamma$, and so $a = 0$. □

*Proof of Theorem 4.2.* We can choose $X^{(i)}$ and the family $\left(X_{k,p}^{(i)}\right)_{0\leq k, 0<p<1}$ in such a way that

$$\mathbb{E}\left[\left|X_{k,p}^{(i)} - X^{(i)}\right|\right] = d_1(G_{k,p}, F),$$

and we can also choose the families $\left(X^{(i)}, (X_{k,p}^{(i)})_{k\geq 0}\right)_{0\leq i\leq I}$ to be i.i.d. Then

$$\begin{aligned} d_1(G_{n,p}, F) &\leq \mathbb{E}\left[\left|\sum_{i=1}^{I} A_{n,p}^{(i)} X_{Z_{n,p}^{(i)},p}^{(i)} + T_{n,p} - \sum_{i=1}^{I} A^{(i)} X^{(i)} - T\right|\right] \\ &\leq \sum_{k=0}^{n-1} \mathbb{E}\left[\sum_{i=1}^{I} \left|A_{n,p}^{(i)} \mathbb{1}_{Z_{n,p}^{(i)}=k}\right|\right] \mathbb{E}\left[\left|X_{k,p}^{(i)} - X^{(i)}\right|\right] + b_{n,p} \\ &\leq \sum_{k=0}^{n-1} \gamma_{k,n,p}\, d_1(G_{k,p}, F) + b_{n,p} \end{aligned}$$

with

$$\begin{aligned} b_{n,p} &= \sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)} - A^{(i)}\right| X^{(i)}\right] + \mathbb{E}\left[\left|T_{n,p} - T\right|\right], \\ \gamma_{k,n,p} &= \sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)}\right| \mathbb{1}_{Z_{n,p}^{(i)}=k}\right]. \end{aligned}$$

Let $M$ be a bound for $\left(\mathbb{E}\left[X_{n,p}\right]\right)_{n,p}$, and set

$$a_{n,p} = d_1(G_{n,p}, F).$$

Let us check the assumptions of Lemma 4.3:

$$0 \leq a_{n,p} \leq \mathbb{E}\left[X_{n,p}\right] + \mathbb{E}\left[X\right] \leq M + \mathbb{E}\left[X\right] \ ;$$

for the second assumption of Lemma 4.3,

$$\limsup_{\mathcal{F}} \sum_{k=0}^{n-1} \gamma_{k,n,p} = \limsup_{\mathcal{F}} \sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)}\right|\right] = \sum_{i=1}^{I} \mathbb{E}\left[\left|A^{(i)}\right|\right] < 1 \ ;$$

$\lim_{\mathcal{F}} b_{n,p} = 0$ by assumption iii), as

$$\sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)} - A^{(i)}\right| X^{(i)}\right] = \mathbb{E}\left[X\right] \sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)} - A^{(i)}\right|\right] \ ;$$

finally

$$\sum_{k \text{ s.t. } (k,p)\notin V} \gamma_{k,n,p} = \sum_{i=1}^{I} \mathbb{E}\left[\left|A_{n,p}^{(i)}\right| \ ; \ (Z_{n,p}^{(i)}, p) \notin V\right].$$

Therefore $d_1(G_{n,p}, F)$ vanishes along $\mathcal{F}$ and the proof of the theorem is now complete. □

The following Theorem is folklore. It gives the means and variances in Theorems 2.1 and 2.2, after some computations.

**Theorem 4.4.** *Suppose that* (19) *holds, where* $\sum_i \mathbb{E}\left[|A^{(i)}|\right] < 1$ *and* $\mathbb{E}\left[|X|\right] < \infty$*; in other words,* $\mathcal{L}(X) = F$*, where* $F$ *is the unique solution in* $D$ *to* $\Psi(F) = F$*. Then*

$$\mathbb{E}[X] = \frac{\mathbb{E}[T]}{1 - \sum_i \mathbb{E}\left[A^{(i)}\right]}. \tag{21}$$

*Moreover, if further* $\sum_i \mathbb{E}\left[|A^{(i)}|^2\right] < 1$ *and* $\mathbb{E}\left[T^2\right] < \infty$*, then* $\mathbb{E}\left[X^2\right] < \infty$ *and*

$$\operatorname{Var}(X) = \frac{\mathbb{E}\left[T^2\right] + 2\,\mathbb{E}[X]\,\mathbb{E}\left[T\sum_i A^{(i)}\right] + \mathbb{E}\left[\left(\sum_i A^{(i)}\right)^2 - 1\right](\mathbb{E}[X])^2}{1 - \sum_i \mathbb{E}\left[A^{(i)\,2}\right]}. \tag{22}$$

*Proof.* Taking expectations in (19) we obtain $\mathbb{E}[X] = \sum_i \mathbb{E}\left[A^{(i)}\right]\mathbb{E}[X] + \mathbb{E}[T]$, which yields (21).

For the second part, let $D_2 = \{\mu \in D : \int x\,d\mu(x) = \mathbb{E}[X], \int x^2\,d\mu(x) < \infty\}$. It is easy to see that now $\Psi$ is a strict contraction in $D_2$ with the $d_2$ metric; hence $\Psi$ has a unique fixed point in $D_2$. Since $D_2 \subset D$, this fixed point must be $F$, which shows that $\mathbb{E}\left[X^2\right] < \infty$. If we square (19) and take the expectation, we obtain

$$\begin{aligned}\mathbb{E}\left[X^2\right] = \mathbb{E}\left[\sum_{i=1}^{I} A^{(i)\,2}\right]\mathbb{E}\left[X^2\right] + \sum_{1\le i\ne j\le I}\mathbb{E}\left[A^{(i)}A^{(j)}\right](\mathbb{E}[X])^2 \\ + 2\sum_{i=1}^{I}\mathbb{E}\left[A^{(i)}T\right]\mathbb{E}[X] + \mathbb{E}\left[T^2\right],\end{aligned}$$

which yields (22). □

## 5. Proofs of Theorems 2.1 and 2.2

We apply Theorem 4.2 to the distributional identity (12), with $I = 2$,

$$\left(A^{(1)}_{n,p}, Z^{(1)}_{n,p}\right) = \left(A_{n,p}, Z_{n,p} - 1\right),$$

and

$$\left(A^{(2)}_{n,p}, Z^{(2)}_{n,p}\right) = \left(B_{n,p}, n - Z_{n,p}\right).$$

Here the distribution of $\left(A^{(i)}_{n,p}, Z^{(i)}_{n,p}\right)$ does not depend on $i$. We verify the assumptions ii)–iv) of Theorem 4.2 for Theorems 2.1 and 2.2 together; for the second theorem take $c = 0$. The first assumption holds true by Proposition 3.5.

**Verification of the second point.** We have

$$\begin{aligned}A^{(1)} &= A = [(1-2c)U + c]^2,\\ A^{(2)} &= B = [(2c-1)U + 1 - c]^2,\end{aligned}$$

and $c \in [0,1]$. Easy computations give

$$\mathbb{E}\left[[(1-2c)U + c]^2\right] + \mathbb{E}\left[[(2c-1)U + 1 - c]^2\right] = \frac{2}{3}(1 - c + c^2) \le \frac{2}{3}.$$

**Verification of the third point.** We must prove the convergence of $A_{n,p}$, $B_{n,p}$ and $T_{n,p}$ to $A$, $B$ and $T(c,U)$, in $L^1$. Recall (13)–(17).

From Proposition 3.2 we know that, conditioned on $U$, $s_\ell \sim \mathrm{Bi}(\lceil nU \rceil - 1, p)$ and thus

$$\mathbb{E}\big((s_\ell - (\lceil nU\rceil - 1)p)^2 \mid U\big) = (\lceil nU\rceil - 1)p(1-p) \le np.$$

Hence, taking the expectation,

$$\mathbb{E}\big(s_\ell - (\lceil nU\rceil - 1)p\big)^2 \le np$$

and thus

$$\|s_\ell - nUp\|_2 \le \|s_\ell - (\lceil nU\rceil - 1)p\|_2 + p \le (np)^{1/2} + p \le 2(np)^{1/2}.$$

Consequently,

$$\left\|\frac{s_\ell}{n} - Uc\right\|_2 \le \left\|\frac{s_\ell}{n} - Up\right\|_2 + |p-c| \to 0, \tag{23}$$

and, similarly but more sharply,

$$\left\|\frac{s_\ell}{n\sqrt{p}} - U\sqrt{c}\right\|_2 \le \frac{2}{\sqrt{n}} + |\sqrt{p} - \sqrt{c}| \to 0. \tag{24}$$

Similarly,

$$\left\|\frac{s_r}{n} - (1-U)c\right\|_2 \to 0 \tag{25}$$

and

$$\left\|\frac{s_r}{n\sqrt{p}} - (1-U)\sqrt{c}\right\|_2 \to 0. \tag{26}$$

From (13), (23) and (25) follows

$$\left\|\frac{Z_{n,p}-1}{n} - \big(U - Uc + (1-U)c\big)\right\|_2 \to 0. \tag{27}$$

It follows easily from Cauchy–Schwarz's inequality that multiplication is a continuous bilinear map $L^2 \times L^2 \to L^1$. Hence (27) yields

$$\|A_{n,p} - A\|_1 = \left\|\left(\frac{Z_{n,p}-1}{n}\right)^2 - \big(U - Uc + (1-U)c\big)^2\right\|_1 \to 0,$$

verifying the first assertion. (27) similarly implies $\|B_{n,p} - B\|_1 \to 0$ too.

For $T_{n,p}$ we first observe that, similarly, from (24) and (26),

$$\left\|\frac{s_\ell s_r}{n^2 p} - U(1-U)c\right\|_1 \to 0.$$

Moreover, since $np \to \infty$, (23) and (25) imply $\left\|s_\ell/n^2p\right\|_1 \le \left\|s_\ell/n^2p\right\|_2 \to 0$ and $\left\|s_r/n^2p\right\|_1 \to 0$.

For the terms $W_{\vec{\ell}}$ and $W_{\vec{r}}$ we use Proposition 3.1. We have $\|S_m - mp\|_2 = \sqrt{mp(1-p)}$ and thus, uniformly for $0 \le m \le n$,

$$\left\|\frac{S_m}{n\sqrt{p}} - \frac{m}{n}\sqrt{c}\right\|_2 \le \frac{1}{\sqrt{n}} + |\sqrt{p} - \sqrt{c}| \to 0,$$

which, using Cauchy–Schwarz again, yields

$$\left\|\frac{S_m(S_m+1)}{2n^2p} - \frac{c}{2}\left(\frac{m}{n}\right)^2\right\|_1 \to 0. \tag{28}$$

Moreover, let $W'_m = \sum_{i=1}^m iY_i$. Then $\mathbb{E}W'_m = \frac{m(m+1)p}{2}$ and

$$\|W'_m - \mathbb{E}W'_m\|_2^2 = \mathrm{Var}\,(W'_m) = \sum_{i=1}^m i^2 p(1-p) \leq m^3 p,$$

and thus

$$\left\| \frac{W'_m}{n^2 p} - \frac{1}{2}\left(\frac{m}{n}\right)^2 \right\|_2 \leq \frac{1}{\sqrt{np}} + \frac{1}{2n} \to 0. \tag{29}$$

Proposition 3.1 now yields, by (28) and (29), uniformly for $m \leq n$,

$$\left\| \frac{W_m}{n^2 p} - \frac{1-c}{2}\left(\frac{m}{n}\right)^2 \right\|_1 \to 0.$$

Consequently, using Proposition 3.2,

$$\left\| \frac{W_{\vec{\ell}}}{n^2 p} - \frac{1-c}{2} U^2 \right\|_1 \to 0,$$
$$\left\| \frac{W_{\vec{r}}}{n^2 p} - \frac{1-c}{2} (1-U)^2 \right\|_1 \to 0.$$

Collecting the various terms above, we find $\|T_{n,p} - T\|_1 \to 0$.

**Verification of the fourth point.** As already noticed at the beginning of the Section, the distribution of $\left(A_{n,p}^{(j)}, Z_{n,p}^{(j)}\right)$ does not depend on $j \in \{1,2\}$, so in order to prove the two theorems, we only have to check that the fourth assumption holds for $j = 1$, for an arbitrary set in each of the two filters:

$$\lim_{\mathcal{F}_i} \mathbb{E}\left[\left|A_{n,p}\right| \,;\, (Z_{n,p} - 1, p) \notin V\right] = 0, \qquad \forall V \in \mathcal{F}_i, \forall i \in \{1,2\}; \tag{30}$$

also, the expectation on the left hand side of (30) is decreasing in $V$, so we need only to check (30) for typical elements of the filters' basis. But for $(n,p) \in V_{N,\varepsilon}$ (resp. for $(n,p) \in \widetilde{V}_{N,\varepsilon}$),

$$\mathbb{E}\left[\left|A_{n,p}\right| \,;\, (Z_{n,p} - 1, p) \notin V_{N,\varepsilon}\right] \leq \left(\frac{N-1}{n}\right)^2,$$
$$\mathbb{E}\left[\left|A_{n,p}\right| \,;\, (Z_{n,p}, p) \notin \widetilde{V}_{N,\varepsilon}\right] \leq \left(\frac{N}{np}\right)^2.$$

## 6. Proofs of Theorems 2.4 and 2.6

The proof of these theorems is done in four steps:

(i) We prove that $X(\lambda)$ defined at (4) is almost surely finite, and has exponentially decreasing distribution tail. Thus it has moments of all orders.
(ii) With the help of a Poisson point process representation of Quicksort, we prove the convergence of certain copies of $X_{n,p}$ to a copy of $X(\lambda)$ for the norm $\|\cdot\|_1$. This entails the weak convergence.
(iii) We prove that $X(\lambda)$ satisfies the functional equation (5), and that (5) has a unique solution under the extra assumptions in Theorem 2.6.
(iv) We compute the first and second moments of $X(\lambda)$, as required for the proof of Theorem 2.4, and we also give an induction formula for moments of larger order.

**Some properties of $X(\lambda)$.** In this Section, we prove some properties of the family of random variables $(X(\lambda))_{\lambda>0}$ defined by (4). Recall that the increasing sequence $(Y_{k,j})_{0\le j\le 2^k}$, defined by the recurrence relation (3), splits $[0,1]$ in $2^k$ intervals, obtained recursively by breaking each of the $2^{k-1}$ intervals of the previous step into two random pieces. For $k \ge 0$ and $1 \le i \le 2^k$, let

$$\begin{aligned} w_{k,i} &= Y_{k,i} - Y_{k,i-1}, \\ M_k &= \max\left\{w_{k,i} : 1 \le i \le 2^k\right\}, \\ F_{k,\alpha} &= \left(\frac{1+\alpha}{2}\right)^k \sum_{1\le i\le 2^k} w_{k,i}^\alpha \\ \mathcal{F}_k &= \sigma\left(Y_{i,j} : i \le k, 1 \le j \le 2^i - 1\right) \\ \mathcal{F} &= (\mathcal{F}_k)_{k\ge 0}\,. \end{aligned}$$

We begin with a simple estimate (see also [7]):

**Lemma 6.1.** $\mathbb{E}\left[w_{k,j}^2\right] = 3^{-k}$.

*Proof.* The length $w_{k,j} = |I_{k,j}|$ is the product of $k$ independent random variables, each uniform on $[0,1]$. Hence $\mathbb{E}\left[w_{k,j}^2\right] = \left(\mathbb{E}\left[U^2\right]\right)^k = 3^{-k}$. $\square$

**Lemma 6.2.** *For $\alpha > 0$, $(F_{k,\alpha})_{k\ge 0}$ is a $\mathcal{F}$-martingale, and $\mathbb{E}\left[F_{k,\alpha}\right] = 1$.*

*Proof.* Clearly $\mathbb{E}\left[F_{0,\alpha}\right] = 1$. Also:

$$\begin{aligned} \mathbb{E}\left[F_{k+1,\alpha}|\mathcal{F}_k\right] &= \left(\frac{1+\alpha}{2}\right)^{k+1} \sum_{i=1}^{2^k} \mathbb{E}\left[w_{k+1,2i-1}^\alpha + w_{k+1,2i}^\alpha|\mathcal{F}_k\right] \\ &= \left(\frac{1+\alpha}{2}\right)^{k+1} \sum_{i=1}^{2^k} w_{k,i}^\alpha \mathbb{E}\left[U_{k,i}^\alpha + (1-U_{k,i})^\alpha\right] \\ &= \left(\frac{1+\alpha}{2}\right)^{k} \sum_{i=1}^{2^k} w_{k,i}^\alpha. \end{aligned}$$

$\square$

Let $\rho = 0.792977\ldots$ denote the larger real solution of the equation $\rho^{-1} = -2e\ln\rho$. Lemma 6.2 entails that

**Lemma 6.3.** $\mathbb{E}\left[M_k\right] \le \rho^k$.

*Proof.* Clearly,

$$M_k^\alpha \le \left(\frac{2}{1+\alpha}\right)^k F_{k,\alpha};$$

thus, for $\alpha \ge 1$,

$$\mathbb{E}\left[M_k\right] \le \left(\mathbb{E}\left[M_k^\alpha\right]\right)^{1/\alpha} \le \left(\frac{2}{1+\alpha}\right)^{k/\alpha} \left(\mathbb{E}\left[F_{k,\alpha}\right]\right)^{1/\alpha} = \left(\frac{2}{1+\alpha}\right)^{k/\alpha}.$$

The rate $\left(\frac{2}{1+\alpha}\right)^{1/\alpha}$ reaches its minimum for $1+\alpha = 4.311\ldots$, a constant that is an old friend of Quicksort and binary search trees [3]. This leads to the desired value for $\rho$. $\square$

A weaker form of this inequality (for $\alpha = 2$), actually sufficient for our purposes, is given in [7]. The sequence $(F_{k,\alpha})_{k\geq 0}$ is a specialization of martingales that are of a great use for the study of general branching random walks, see for instance [1], of which binary search trees are a special case [11, 12].

**Lemma 6.4.** $\mathbb{E}\left[X(\lambda)\right] = 1$.

*Proof.* Set $\mathcal{F}_\infty = \sigma\left(Y_{k,j}, k \geq 0, 1 \leq j \leq 2^k - 1\right)$. Inspecting (7), we see that

$$\mathbb{E}\left[X(\lambda) \mid \mathcal{F}_\infty\right] = \frac{1}{2} \sum_{k\geq 1} \left(\frac{2}{3}\right)^k F_{k,2},$$

because, conditionally given $\mathcal{F}_\infty$, the expected number of points of $\Pi_{k,j}$ is $\lambda w_{k,j}$ and each of them has an expected contribution $w_{k,j}/(2\lambda)$ to $X(\lambda)$. □

As a consequence of Lemma 6.3, we have

**Theorem 6.5.** *For each fixed $\lambda > 0$, the distribution tail $\mathbb{P}\left(X(\lambda) \geq x\right)$ decreases exponentially fast.*

*Proof.* Equivalently, we prove this result for $\Xi(\lambda) = \lambda X(\lambda)$. Since

$$\left|x - Y_{k,J_k(x)}\right| \leq M_k,$$

we have

$$\Xi(\lambda) \leq \sum_{(k,x)\in\Pi} M_k = \sum_{k\geq 1} N_k M_k,$$

where $N_k = |\Pi_k|$ is a Poisson random variable with mean $\lambda$. We split the tail of this bound on $\Xi(\lambda)$ as follows:

$$\begin{aligned}\mathbb{P}\left(\Xi(\lambda) \geq x\right) &\leq \mathbb{P}\left(\sum_{k\geq 1} N_k M_k \geq x\right)\\ &\leq p_1 + p_2,\end{aligned}$$

in which

$$\begin{aligned}p_1 &= \mathbb{P}\left(\sum_{1\leq k\leq m} N_k M_k \geq x/2\right),\\ p_2 &= \mathbb{P}\left(\sum_{k>m} N_k M_k \geq x/2\right).\end{aligned}$$

We have, by the standard Chernoff bound for the Poisson distribution,

$$p_1 \leq \mathbb{P}\left(\sum_{1\leq k\leq m} N_k \geq x/2\right) \leq \exp\left(\frac{x}{2}\left(1 - \ln(x/2m\lambda)\right) - n\lambda\right),$$

the last inequality holding only for $m \leq \frac{x}{2\lambda}$. Also

$$\begin{aligned}p_2 &\leq \mathbb{P}\left(\sum_{k>m} N_k \rho^{k/2} \geq x/2\right) + \mathbb{P}\left(\exists k > m :\, M_k > \rho^{k/2}\right)\\ &\leq \left(1 + \frac{2\lambda}{x}\right) \frac{\rho^{(m+1)/2}}{1 - \sqrt{\rho}},\end{aligned}$$

using a Markov first moment inequality to bound both terms. For any $\alpha$ in $(0,1)$, the choice $m \sim \frac{\alpha x}{2\lambda}$ leads to an exponential decrease of the tail. $\square$

**Convergence of $X_{n,p}$ to $X(\lambda)$.** We assume that the input list for Quicksort contains the integers $\{1,2,\dots,n\}$ in random order. We model our error-prone Quicksort as follows using the variables $U_{k,j}$ and $\Pi$ in Section 2, but with the intensity $\lambda$ of $\Pi$ replaced by $\lambda(n,p) = -n\ln(1-p)$:

In the first step, we use the pivot $p_{1,1} = \lceil nU_{0,1}\rceil$ and let for each $i$ (except the pivot) there be an error in the comparison of $i$ and the pivot if $\Pi_1 \cap (\frac{i-1}{n}, \frac{i}{n}] \neq \emptyset$. (Recall that $\Pi_k = \{x : (k,x) \in \Pi\}$.) Note that our choice of $\lambda(n,p)$ yields the right error probability $p$.

Let $p'_{1,1}$ be the position of the pivot after the first step. (This position was earlier denoted $Z_{n,p}$; it may differ from $p_{1,1}$ because of errors.) The items of the left sublist will thus be placed in positions $1,\dots,p'_{1,1}-1$ and those in the right sublist in positions $p'_{1,1}+1,\dots,n$. Let $p'_{1,0}=0$ and $p'_{1,2}=1+n$.

When the $k$-th step begins, we have a set of $2^{k-1}$ sublists $(\ell_{k-1,j})_{j=1,\dots,2^{k-1}}$, the elements of $\ell_{k-1,j}$ being in positions $p'_{k-1,j-1}+1,\dots,p'_{k-1,j}-1$, $j=1,\dots,2^{k-1}$ (with the convention that the sublist is empty when $p'_{k-1,j}-p'_{k-1,j-1} \leq 1$). In each nonempty such sublist we choose as pivot the item with rank $\lceil U_{k-1,j}(p'_{k-1,j} - p'_{k-1,j-1}-1)\rceil$, in this sublist, so that its position in the final output will be exactly

$$p_{k,2j-1} = p'_{k-1,j-1} + \lceil U_{k-1,j}(p'_{k-1,j} - p'_{k-1,j-1}-1)\rceil, \tag{31}$$

in case no errors occurs while processing the sublist. We assume an error is made when comparing the element at position $i$ with the pivot $p_{k,2j-1}$ if $\Pi_k \cap (\frac{i-1}{n},\frac{i}{n}] \neq \emptyset$. Let $p'_{k,2j} = p'_{k-1,j}$. Let $p'_{k,2j-1}$ be the position of the pivot $p_{k,2j-1}$ after the comparisons (as in the first step, $p'_{k,2j-1}$ may differ from $p_{k,2j-1}$ because of errors); let $p'_{k,2j-1} = p'_{k-1,j}$ if the sublist was empty. Set

$$y_{k,j} = p_{k,j}/n \quad \text{and} \quad y'_{k,j} = p'_{k,j}/n.$$

We expect $y_{k,j}$ and $y'_{k,j}$ to converge to $Y_{k,j}$ as $n \to +\infty$.

This procedure (stopped when there are no more nonempty sublists) is an exact simulation of the erratic Quicksort, so we may assume that $I(n,p)$ is the number of inversions created by it. As in Section 3, let $I^{(k)}(n,p)$ be the number of inversions created at step $k$, so

$$I(n,p) = \sum_{k=1}^{\infty} I^{(k)}(n,p).$$

We will prove that, using the notation of (7),

$$\delta_{n,k} = \left\| \frac{1}{n^2 p} I^{(k)}(n,p) - \frac{1}{\lambda(n,p)} \sum_{j=1}^{2^k} \sum_{x\in\Pi_{k,j}} |x - x_{k,j}| \right\|_1 \longrightarrow 0 \tag{32}$$

for each $k$. Since also, by Lemmas 3.4 and 6.1,

$$\begin{aligned}\delta_{n,k} &\le \frac{1}{n^2 p}\mathbb{E}\left[I^{(k)}(n,p)\right] + \frac{1}{\lambda(n,p)}\mathbb{E}\left[\sum_{j=1}^{2^k}\sum_{x\in\Pi_{k,j}}|x-x_{k,j}|\right] \\ &= \frac{1}{n^2 p}\mathbb{E}\left[I^{(k)}(n,p)\right] + \mathbb{E}\left[\frac{1}{2}\sum_{j=1}^{2^k} w_{k,j}^2\right] \le \left(\frac{2}{3}\right)^k,\end{aligned}$$

it follows by dominated convergence that, using (7),

$$\left\|X_{n,p} - X\big(\lambda(n,p)\big)\right\|_1 \le \sum_{k=1}^{\infty}\delta_{n,k} \longrightarrow 0.$$

Moreover, $\lambda(n,p)\to\lambda$, and it follows easily from (8) that $\left\|X\big(\lambda(n,p)\big) - X(\lambda)\right\|_1 \to 0$. Hence we have $\mathbb{E}\,|X_{n,p} - X(\lambda)| \to 0$, which proves the convergence.

It remains to verify (32). Set

$$X^{(k)} = \sum_{j=1}^{2^k}\sum_{x\in\Pi_{k,j}}|x-x_{k,j}|.$$

Relation (32) is equivalent to

$$\mathbb{E}\left|\frac{1}{n}I^{(k)}(n,p) - X^{(k)}\right| \to 0. \tag{33}$$

For simplicity, we write in the sequel $\lambda$ instead of $\lambda(n,p)$. We begin with a lemma.

**Lemma 6.6.** *For each $k$ and $j$,*

$$\max\left\{\left\|Y_{k,j} - y'_{k,j}\right\|_1, \left\|Y_{k,j} - y_{k,j}\right\|_1\right\} \le \frac{k(1+\lambda)}{n}.$$

*Proof.* Recall that $p'_{k,j} = ny'_{k,j}$, so (31) translates to

$$p_{k,2j-1} = ny'_{k-1,j-1} + \lceil U_{k-1,j}(ny'_{k-1,j} - ny'_{k-1,j-1} - 1)\rceil,$$

We use induction on $k$. Comparing the definitions of $Y_{k,j}$ and $y'_{k,j}$, we see that it suffices to consider an odd $j = 2l-1$, and in that case there are three sources of a difference:

(i) The differences between $y'_{k-1,l-1}$ and $Y_{k-1,l-1}$ and between $y'_{k-1,l}$ and $Y_{k-1,l}$. By the induction hypothesis, this contributes at most $(k-1)(1+\lambda)/n$.

(ii) The $-1$ inside (and the rounding by) the ceiling function. This contributes at most $1/n$.

(iii) The shift of the pivot, from $p_{k,2j-1}$ to $p'_{k,2j-1}$, caused by the erroneous comparisons. The shift is bounded by the total number of errors at step $k$, so its mean is less than $\lambda$, and the contribution is less than $\lambda/n$.

□

We return to proving (33). For $k=1$, $I^{(1)}(n,p)$ is just $t(n,p)$ studied in Section 3, and (10) yields

$$I^{(1)}(n,p) = s_\ell s_r + s_\ell + s_r + W_{\vec{\ell}} + W_{\vec{r}}.$$

Let $E_1$ be the set of items $i$ such that an error was made in the comparison with $p_{1,1}$. Relation (11) entails that

$$\sum_{i\in E_1} |i - p_{1,1}| = W_{\vec{\ell}} + W_{\vec{r}} + \tfrac{1}{2}s_\ell(s_\ell+1) + \tfrac{1}{2}s_r(s_r+1).$$

We shall denote this last sum $\tilde{I}^{(1)}(n,p)$. Thus, we have

$$\left|I^{(1)}(n,p) - \tilde{I}^{(1)}(n,p)\right| = \left|s_\ell s_r + s_\ell + s_r - \tfrac{1}{2}s_\ell(s_\ell+1) - \tfrac{1}{2}s_r(s_r+1)\right| \le s_\ell^2 + s_r^2.$$

Furthermore

$$\mathbb{E}\left[s_\ell^2 \mid \lceil nU_{1,1}\rceil = m\right] = (m-1)p(1-p) + ((m-1)p)^2 \le np + n^2p^2. \tag{34}$$

Hence,

$$\left\|I^{(1)}(n,p) - \tilde{I}^{(1)}(n,p)\right\|_1 = \mathcal{O}(1).$$

Moreover, $\frac{1}{n}\tilde{I}^{(1)}(n,p) = \sum_{i\in E_1} |\frac{i}{n} - y_{1,1}|$ differs from $X^{(1)} = \sum_{j=1}^2 \sum_{x\in\Pi_{1,j}} |x - x_{1,j}|$ in (33) in four ways only (recall that $x_{1,1} = x_{1,2} = Y_{1,1}$):

(i) $i/n$ differs from $x$ by at most $1/n$. Since the expected number of terms is not larger than $\lambda$, this gives a contribution $\mathcal{O}(1/n)$.

(ii) $|y_{1,1} - x_{1,j}| = |y_{1,1} - Y_{1,1}|$, which by Lemma 6.6 has expectation $\mathcal{O}(1/n)$. Thus this too gives a contribution $\mathcal{O}(1/n)$.

(iii) If there are two or more points in $\Pi_1 \cap (\frac{i-1}{n}, \frac{i}{n}]$ for some $i$, $X^{(1)}$ contains more terms than $\frac{1}{n}\tilde{I}^{(1)}(n,p)$. It is easily seen that the expected number of such extra points in each interval $(\frac{i-1}{n}, \frac{i}{n}]$ is less than $(\lambda/n)^2$, and each point contributes at most 1 to $X^{(1)}$.

(iv) Each point in $\Pi_1 \cap (\frac{p_{1,1}-1}{n}, \frac{p_{1,1}}{n}]$ contributes for an extra term in $X^{(1)}$ again. The expected number of such extra points is $\lambda/n$ and each of these terms contributes at most 1 to $X^{(1)}$.

This verifies (33) for $k=1$.

For $k \ge 2$ we argue similarly. We can approximate $I^{(k)}(n,p)$ by the sum of the distances between the errors and the respective pivots,

$$\tilde{I}^{(k)}(n,p) = \sum_{j\le 2^{k-1}} \sum_{i\in E_{k-1,j}} |i - p_{k,2j-1}|,$$

as follows: Let $E_{k,j}$ be the set of items $i \in \ell_{k,j}$ subject to error when compared with $p_{k+1,2j-1}$, and let $\mathcal{G}_k$ be the $\sigma$-algebra generated by $(U_{\ell,j})_{\ell\le k, j\le 2^\ell}$ and $\Pi_1 \cup \Pi_2 \cup \dots \cup \Pi_{k-1}$. As for $k=1$, using relation (11), we obtain the following bound:

$$\begin{aligned}\mathbb{E}\left[\left|I^{(k)}(n,p) - \tilde{I}^{(k)}(n,p)\right| \;\middle|\; \mathcal{G}_k\right] &\le \sum_{j\le 2^{k-1}} \left(p^2\left(\#\ell_{k-1,j}\right)^2 + p\#\ell_{k-1,j}\right)\\ &\le 2^{k-1}(n^2p^2 + np) = \mathcal{O}(1),\end{aligned}$$

and as a consequence,

$$\left\|I^{(k)}(n,p) - \tilde{I}^{(k)}(n,p)\right\|_1 = \mathcal{O}(1).$$

Now,

$$\frac{1}{n}\tilde{I}^{(k)}(n,p) = \sum_{j\le 2^{k-1}} \sum_{i\in E_{k-1,j}} \left|\frac{i}{n} - y_{k,2j-1}\right|$$

differs from $X^{(k)} = \sum_{j=1}^{2^k} \sum_{x\in\Pi_{k,j}} |x - x_{k,j}|$ in (33) in the same four ways as for $k = 1$, plus an extra fifth way:

(i) See the case $k = 1$.
(ii) $|y_{k,2j-1} - x_{k,2j-1}| = |y_{k,2j-1} - x_{k,2j}| = |y_{k,2j-1} - Y_{k,2j-1}|$, which by Lemma 6.6 has expectation $\mathcal{O}(1/n)$. Thus this too gives a contribution $\mathcal{O}(1/n)$.
(iii) Two or more points in $\Pi_k \cap (\frac{i-1}{n}, \frac{i}{n}]$ for some $i$, see the case $k = 1$.
(iv) Each point in $\Pi_k \cap (-1/n + y_{k,2j-1}, y_{k,2j-1}]$ contributes for an extra term in $X^{(k)}$. The expected number of such extra points is $\lambda 2^{k-1}/n$ and each of these terms contributes at most 1 to $X^{(k)}$.
(v) There is a new source of error in this approximation, because some points $x$ in $\Pi_k$ and the corresponding positions $i = \lceil nx \rceil$ belong to subintervals that do not correspond to each other, because the endpoints $y'_{k-1,j}$ differ somewhat from $Y_{k-1,j}$. By Lemma 6.6, the expected number of such cases is $\mathcal{O}(1/n)$, so again we get a contribution of order $\mathcal{O}(1/n)$ only.

This verifies (33) and thus the convergence of $X_{n,p}$ to $X(\lambda)$.

**The distributional identity for $X(\lambda)$.** We check that $X(\lambda)$ satisfies the distributional identity and some side conditions needed for the computations of moments.

**Proposition 6.7.** *$(X(\lambda))_{\lambda>0}$ is a solution of (5). Moreover, $\mathbb{E}[X(\lambda)^n] < \infty$ and $\lambda^n \mathbb{E}[X(\lambda)^n] \to 0$ as $\lambda \to 0$, for $n \geq 1$.*

*Proof.* All moments are finite by Theorem 6.5. Moreover, $\mathbb{E}[(\lambda X(\lambda))^n] \to 0$ as $\lambda \to 0$ by (8) and dominated convergence.

For $a < b$, let $\Pi(a, b)$ be a Poisson point process of intensity $\lambda$ on $\mathbb{N}^* \times [a, b]$, and let $\{U_{k,j} : k \geq 0, 1 \leq i \leq 2^k\}$ be independent uniform random variables as in Section 2, and further independent of $\Pi(a, b)$. Define $\{Y_{k,j} : k \geq 0, 1 \leq i \leq 2^k\}$ and $J_k(x)$ as in Section 2, with the slight modification

$$Y_{0,0} = a \qquad \text{and} \qquad Y_{0,1} = b$$

and set

$$X(\lambda, a, b) = \frac{1}{\lambda} \sum_{(k,x)\in\Pi(a,b)} \left|x - Y_{k,J_k(x)}\right|.$$

Note that $X(\lambda, 0, 1) = X(\lambda)$. Shifting and rescaling $\Pi(a, b)$, we obtain

$$X(\lambda, a, b) \overset{\text{law}}{=} X(\lambda, 0, b - a) \overset{\text{law}}{=} (b - a)^2 X(\lambda(b - a)).$$

Let us split $X(\lambda)$: we have

$$\begin{aligned}
X(\lambda) &= X_0(\lambda) + X_1(\lambda) + X_2(\lambda)\\
\lambda X_0(\lambda) &= \sum_{(1,x)\in\Pi(0,1)} \left|x - Y_{1,1}\right|\\
\lambda X_1(\lambda) &= \sum_{\substack{(k,x)\in\Pi(0,1)\\ k\geq 2, x\leq Y_{1,1}}} \left|x - Y_{k,J_k(x)}\right|,\\
\lambda X_2(\lambda) &= \sum_{\substack{(k,x)\in\Pi(0,1)\\ k\geq 2, x\geq Y_{1,1}}} \left|x - Y_{k,J_k(x)}\right|.
\end{aligned}$$

We see, using general properties of Poisson point processes and the recursive construction of $\{Y_{k,j} : k \geq 0, 1 \leq i \leq 2^k\}$, that

$$
\begin{aligned}
(X_0(\lambda), X_1(\lambda), X_2(\lambda)) &\stackrel{\text{law}}{=} (\Theta(\lambda, Y_{1,1}), X(\lambda, 0, Y_{1,1}), \widetilde{X}(\lambda, Y_{1,1}, 1)) \\
&\stackrel{\text{law}}{=} (\Theta(\lambda, Y_{1,1}), Y_{1,1}^2 X(\lambda Y_{1,1}), (1 - Y_{1,1})^2 \widetilde{X}(\lambda(1 - Y_{1,1}))),
\end{aligned}
$$

in the sense that, conditionally given that $Y_{1,1} = u$, $X_0(\lambda)$, $X_1(\lambda)$ and $X_2(\lambda)$ are independent and distributed as $\Theta(\lambda, u)$, $u^2 X(\lambda u)$, $(1-u)^2 X(\lambda(1-u))$, respectively. Also $Y_{1,1} = U_{0,1}$ is uniformly distributed on $[0, 1]$. □

**Uniqueness of solutions of** (5)**.** Let $\mu = (\mu_\lambda)_{\lambda>0}$ and $\theta = (\theta_\lambda)_{\lambda>0}$ be two solutions of (5) in $\mathcal{M}$. Let $Y = (Y(\lambda))_{\lambda>0}$ and $Z = (Z(\lambda))_{\lambda>0}$ denote two measurable processes representing respectively $\mu$ and $\theta$, in the sense of Remark 2.5 (i). Without loss of generality, we can assume that $Y$ and $Z$ share the same underlying probabilistic space, and the same exponent $\alpha$. Then, by definition of $\mathcal{M}$, for $\Lambda > 0$,

$$
d_\Lambda(Y, Z) = \sup_{(0,\Lambda)} \mathbb{E}\left[\lambda^\alpha |Y(\lambda) - Z(\lambda)|\right]
$$

is finite. Let $\delta$ denote the infimum of $d_\Lambda(\hat{Y}, \hat{Z})$ over all couples of representations $(\hat{Y}, \hat{Z})$ of $\mu$ and $\theta$, lying on the same probabilistic space, and assume that $\delta > 0$. Let $(Y_0, Z_0)$ be such a couple of representations, satisfying furthermore

$$
d_\Lambda(Y_0, Z_0) < \delta\ \frac{3 - \alpha}{2}.
$$

Consider a probabilistic space on which are defined three *independent* random variables $(Y_1, Z_1)$, $(Y_2, Z_2)$ and $U$, $(Y_1, Z_1)$ and $(Y_2, Z_2)$ being two copies of $(Y_0, Z_0)$, $U$ being uniform on $(0, 1)$. Finally, for every $\lambda > 0$, set

$$
\begin{aligned}
\hat{Y}(\lambda) &= U^2 Y_1(\lambda U) + (1 - U)^2 Y_2(\lambda(1 - U)) + \Theta(\lambda, U), \\
\hat{Z}(\lambda) &= U^2 Z_1(\lambda U) + (1 - U)^2 Z_2(\lambda(1 - U)) + \Theta(\lambda, U).
\end{aligned}
$$

Then $\hat{Y}$ and $\hat{Z}$ are representations of $\mu$ (resp. $\theta$) and satisfy Remark 2.5 (ii). Moreover, we have, for $\lambda \in (0, \Lambda)$,

$$
\begin{aligned}
\mathbb{E}\left[\lambda^\alpha \left|\hat{Y}(\lambda) - \hat{Z}(\lambda)\right|\right] &= \mathbb{E}\big[\lambda^\alpha \big|U^2 Y_1(\lambda U) + (1 - U)^2 Y_2(\lambda(1 - U)) \\
&\qquad\qquad - U^2 Z_1(\lambda U) - (1 - U)^2 Z_2(\lambda(1 - U))\big|\big] \\
&\leq 2\mathbb{E}\left[\lambda^\alpha U^2 \left|Y_1(\lambda U) - Z_1(\lambda U)\right|\right] \\
&\leq 2 \int_0^1 u^{2-\alpha}\ \mathbb{E}\left[(\lambda u)^\alpha \left|Y_1(\lambda u) - Z_1(\lambda u)\right|\right] du \\
&< \delta,
\end{aligned}
$$

leading to a contradiction.

**Moments of** $X(\lambda)$**.** The aim of this Section is the computation of moments of $X(\lambda)$, completing the proof of Theorem 2.4. If one uses directly (5), the computations of moments by induction are hardly tractable because all three terms on the right of (5) depend on $U$. To circumvent this problem, we consider a new distributional identity

$$
W(\lambda) \stackrel{\text{law}}{=} \xi(\lambda) + U W(\lambda U) + (1 - U) \widetilde{W}(\lambda(1 - U)), \tag{35}
$$

in which

- $\xi(\lambda)$ is as in Section 2; equivalently, $\xi(\lambda) = \sum_{x\in\Pi_1} x$;
- $\xi(\lambda)$ and $\left(U, W(\lambda U), \widetilde{W}(\lambda(1-U))\right)$ are independent;
- conditionally, given $U = u$, $W(\lambda U)$ and $\widetilde{W}(\lambda(1-U))$ are independent and distributed as $W(\lambda u)$ and $W(\lambda(1-u))$, respectively.

The next Propositions establish relations between $X(\lambda)$ and solutions of (35), eventually providing an algorithm for the computation of moments of $X(\lambda)$ (see (36) and (39)).

**Proposition 6.8.** *The family* $(Y(\lambda))_{\lambda>0} = (\xi(\lambda) + \lambda X(\lambda))_{\lambda>0}$, *in which* $\xi(\lambda)$ *and* $X(\lambda)$ *are assumed independent, is a solution of* (35).

**Proposition 6.9.** *The $n$-th moment of $Y(\lambda)$ is a polynomial of degree $n$ in the variable $\lambda$.*

Before proving Propositions 6.8 and 6.9, we need a lemma.

**Lemma 6.10.** *The $n$-th moment $g_n(\lambda) = \mathbb{E}\left[\xi(\lambda)^n\right]$ is a polynomial of degree $n$ with nonnegative coefficients and for $n \geq 1$, $g_n(0) = 0$.*

*Proof.* Owing to Campbell's Theorem [9, p.28], we have

$$\mathbb{E}\left[e^{s\xi(\lambda)}\right] = \exp\left(\lambda\left(\mathbb{E}\left[e^{sU}\right] - 1\right)\right) = \exp\left(\lambda\left(\frac{s}{2!} + \frac{s^2}{3!} + \dots\right)\right).$$

Expanding the last expression gives the lemma. □

*Proof of Proposition 6.8.* To show (35), it is enough to show

$$\begin{aligned}\lambda X(\lambda) &\overset{\text{law}}{=} UY(\lambda U) + (1-U)\widetilde{Y}(\lambda(1-U)) \\ &\overset{\text{law}}{=} U\xi(\lambda U) + \lambda U^2 X(\lambda U) + (1-U)\widetilde{\xi}(\lambda(1-U)) + \lambda(1-U)^2\widetilde{X}(\lambda(1-U)),\end{aligned}$$

where, as usual, conditioned on $U = u$, the terms on the right hand side are independent with the right distributions. This follows immediately from (5), since

$$\lambda\Theta(\lambda, u) \overset{\text{law}}{=} u\xi(\lambda u) + (1-u)\widetilde{\xi}(\lambda(1-u)).$$

□

*Proof of Proposition 6.9.* Consider the sequence of integral equations

$$P_0(\lambda) = 1, \qquad P_n(\lambda) = 2\int_0^1 u^n P_n(\lambda u)du + \psi_n(\lambda), \quad n \geq 1, \tag{36}$$

in which

$$\psi_n(\lambda) = \sum_{\substack{r+k+\ell=n \\ k<n,\ell<n}} \binom{n}{r, k, \ell} g_r(\lambda) \int_0^1 u^k(1-u)^\ell P_k(\lambda u)P_\ell(\lambda(1-u))\, du, \tag{37}$$

where $g_r$ is the $r$-th moment of $\xi(\lambda)$. Proposition 6.9 is a consequence of the next lemma. □

**Lemma 6.11.** *The induction formula* (36) *and the initial condition $P_1(0) = 0$ defines a unique sequence of polynomials, $(P_n(\lambda))_{n\geq 0}$. Furthermore, $P_n$ has degree $n$, and vanishes at 0. For $n \geq 1$, the $n$-th moment $\mathbb{E}\left[Y(\lambda)^n\right]$ is equal to $P_n(\lambda)$.*

*Proof.* Consider $n \geq 1$ and assume that the properties in the lemma hold for $1 \leq m \leq n-1$. Then, for $k$ and $\ell$ smaller than $n$, and $r+k+\ell = n$, the expression

$$g_r(\lambda) \int_0^1 u^k (1-u)^\ell P_k(\lambda u) P_\ell(\lambda(1-u))\, du$$

is a polynomial with degree $n$ and, due to Lemma 6.10, vanishes at 0. Thus, in this case, $\psi_n(\lambda)$ is a polynomial with degree $n$, vanishing at 0. It is now easy to check that a polynomial $P_n(\lambda)$ satisfies (36) if and only if, for $(n,k) \neq (1,0)$,

$$\left[\lambda^k\right] P_n = \frac{n+k+1}{n+k-1} \; \left[\lambda^k\right] \psi_n. \tag{38}$$

Also, by the induction assumptions,

$$\begin{aligned}
\mathbb{E}\left[Y(\lambda)^n\right] &= \mathbb{E}\left[\left(\xi(\lambda) + UY(U\lambda) + (1-U)\widetilde{Y}((1-U)\lambda)\right)^n\right] \\
&= \sum_{r+k+\ell=n} \binom{n}{r,k,\ell} g_r(\lambda) \mathbb{E}\left[U^k(1-U)^\ell Y(U\lambda)^k \widetilde{Y}((1-U)\lambda)^\ell\right] \\
&= 2\mathbb{E}\left[U^n Y(\lambda U)^n\right] + \psi_n(\lambda).
\end{aligned}$$

Note that $\psi_n(\lambda) \geq 0$ for $\lambda \geq 0$. By Remark 2.5, $\lambda \to f_n(\lambda) = \mathbb{E}\left[Y(\lambda)^n\right]$ is nonnegative and measurable. Thus, for $\lambda > 0$, we can rewrite the previous equation:

$$\begin{aligned}
f_n(\lambda) &= 2\int_0^1 u^n f_n(\lambda u) du + \psi_n(\lambda) \\
&= 2\lambda^{-n-1} \int_0^\lambda v^n f_n(v) dv + \psi_n(\lambda).
\end{aligned}$$

Since $f_n(\lambda)$ is assumed to be finite and $\psi_n(\lambda) \geq 0$, the integral on the right hand side is convergent, and thus it is a continuous function of $\lambda$. As a consequence $f_n$ belongs to $C^\infty(0,+\infty)$, and is a solution on $(0,+\infty)$ of the following differential equation:

$$\lambda f_n'(\lambda) + (n-1) f_n(\lambda) = (n+1)\psi_n(\lambda) + \lambda \psi_n'(\lambda).$$

by Proposition 6.7 and Lemma 6.10, $\lambda^{n-1} f_n(\lambda) \to 0$ as $\lambda \to 0$, but the general solution of the differential equation is $P_n(\lambda) + C\,\lambda^{-n+1}$. Thus $f_n = P_n$ on $(0,+\infty)$. ☐

As a consequence of these results, we deduce that:

**Proposition 6.12.** *The function* $\lambda \longrightarrow \lambda^n \mathbb{E}\left[X(\lambda)^n\right]$ *is a polynomial of degree* $n$ *that vanishes at* $0$.

*Proof.* Since $Y(\lambda) = \xi(\lambda) + \lambda X(\lambda)$, with independent summands, we obtain

$$P_m(\lambda) = \mathbb{E}\left[Y(\lambda)^m\right] = \sum_{0 \leq k \leq m} \binom{m}{k} \mathbb{E}\left[\xi(\lambda)^{m-k}\right] \lambda^k \mathbb{E}\left[X(\lambda)^k\right]. \tag{39}$$

The result follows by induction. ☐

**Computation of the first moments.** The moments of $Y(\lambda)$, and thus of $X(\lambda)$, can be computed up to arbitrary order with the help of (36) and (39). For the first two moments, the calculations run as follows. Expanding

$$\exp\left(\lambda\left(\frac{s}{2!}+\frac{s^2}{3!}+\dots\right)\right),$$

in the proof of Lemma 6.10 we obtain

**Lemma 6.13.**

$$g_1(\lambda)=\mathbb{E}\left[\xi(\lambda)\right]=\frac{1}{2}\lambda \qquad \textit{and} \qquad g_2(\lambda)=\mathbb{E}\left[\xi(\lambda)^2\right]=\frac{1}{3}\lambda+\frac{1}{4}\lambda^2.$$

**Proposition 6.14.**

$$\lambda\mathbb{E}\left[X(\lambda)\right]=\mathbb{E}\left[\Xi(\lambda)\right]=\lambda \qquad \textit{and} \qquad \lambda^2\mathrm{Var}\left(X(\lambda)\right)=\mathrm{Var}\left(\Xi(\lambda)\right)=\frac{1}{3}\lambda+\frac{1}{12}\lambda^2.$$

*Proof.* Taking $n=1$ in (37) and (38), we find, using Lemma 6.13,

$$\begin{aligned}\psi_1(\lambda)&=g_1(\lambda)=\tfrac{1}{2}\lambda,\\ P_1(\lambda)&=\tfrac{3}{1}\cdot\tfrac{1}{2}\lambda=\tfrac{3}{2}\lambda.\end{aligned}$$

Taking $n=2$, we similarly find

$$\begin{aligned}\psi_2(\lambda)&=g_2(\lambda)+2\cdot 2g_1(\lambda)\int_0^1 uP_1(\lambda u)\,du+2\int_0^1 u(1-u)P_1(\lambda u)P_1(\lambda(1-u))\,du\\ &=\tfrac{1}{3}\lambda+\tfrac{7}{5}\lambda^2,\\ P_2(\lambda)&=\tfrac{4}{2}\cdot\tfrac{1}{3}\lambda+\tfrac{5}{3}\cdot\tfrac{7}{5}\lambda^2=\tfrac{2}{3}\lambda+\tfrac{7}{3}\lambda^2.\end{aligned}$$

Since $Y(\lambda)=\xi(\lambda)+\lambda X(\lambda)$, with independent summands,

$$P_1(\lambda)=\mathbb{E}\left[Y(\lambda)\right]=\mathbb{E}\left[\xi(\lambda)\right]+\lambda\mathbb{E}\left[X(\lambda)\right],$$

which by Lemma 6.13 yields $\lambda\mathbb{E}\left[X(\lambda)\right]=\lambda$. Similarly,

$$\lambda^2\mathbb{E}\left[X(\lambda)^2\right]=P_2(\lambda)-\mathbb{E}\left[\xi(\lambda)^2\right]-2\mathbb{E}\left[\xi(\lambda)\right]\mathbb{E}\left[\lambda X(\lambda)\right]=\tfrac{1}{3}\lambda+\tfrac{13}{12}\lambda^2,$$

which yields the variance formula. □

The formulas for mean and variance of $X(\lambda)$ can also be obtained directly from (8) and Lemma 6.13; we leave this as an exercise.

## 7. Concluding remarks

We have presented a probabilistic analysis of Quicksort when some comparisons can err. Analysing other sorting algorithms such as merge sort, insertion sort or selection is even more intricate. They do not fit into the model presented in this paper and further more involved probabilistic models/arguments are required. We conjecture that the same normalization holds for the number of inversions in the *output of merge sort* for $n=2^m\to+\infty$, $p=\lambda/n$, and that the limit law $\widehat{X}(\lambda)$ satisfies

$$\mathbb{E}\left[\widehat{X}(\lambda)\right]=\sum_{k\geq 0}\frac{2^k}{(2^k+2)(2^k+3)}=0.454674373\dots<\mathbb{E}\left[X(\lambda)\right].$$

## 8. Acknowledgements

A more rigorous formulation and a shorter proof of Theorem 2.6 are born from discussions with Uwe Rösler. Also, we thank two anonymous referees, whose careful reading led to substantial improvements.

## References

[1] J.D. Biggins. Uniform convergence of martingales in the branching random walk, *Ann. Probab.*, 20(1): 137–151, 1992.
[2] N. Bourbaki. *Elements of mathematics, General topology, Part 1,* Addison-Wesley, 1966.
[3] L. Devroye. A note on the height of binary search trees, *J. Assoc. Comput. Mach.*, 33: 489–198, 1986.
[4] S. Cambanis, G. Simon & W. Stout. Inequalities for $Ek(X, Y)$ when the marginals are fixed, *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete*, 36: 285–294, 1976.
[5] W. Feller. *An introduction to probability theory and its applications, Vol. II,* 2nd ed., John Wiley & Sons, 1971.
[6] F. Ergün, S. Kannan, S.R. Kummar, R. Rubinfeld & M. Vishanathan. Spot-Checkers, *J. Comput. Sys. Sci.*, 60(3): 717–751, 2000.
[7] R. Grüber & U. Rösler. Asymptotic distribution theory for Hoare's selection algorithm, *Adv. Appl. Prob.*, 28: 252–269, 1996.
[8] I. Karatzas & S.E. Shreve. *Brownian motion and stochastic calculus,* 2nd ed., Springer-Verlag, 1991.
[9] J.F.C. Kingman. *Poisson processes,* Oxford University Press, New York, 1993.
[10] D.E. Knuth. *The Art of Computer Programming. Vol. 3 : Sorting and Searching,* 2nd ed., John Wiley & Sons, 1986.
[11] H.M. Mahmoud. *Evolution of random search trees,* John Wiley & Sons, 1992.
[12] B. Pittel. On growing random binary trees, *J.Math.Anal.Appl.*, 103: 461–480, 1984.
[13] S.T. Rachev. *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, U.K., 1991.
[14] U. Rösler. A limit theorem for "Quicksort", *Theor. Inf. Appl.*, 25: 85–100, 1991.
[15] U. Rösler & L. Rüschendorf. The contraction method for recursive algorithms, *Algorithmica*, 29: 3-33, 2001.

INRIA-Lorraine and LORIA, Université Henri Poincaré-Nancy I, BP 239, 54506, Vandœuvre-lès-Nancy, France
*E-mail address*: Laurent.Alonso@loria.fr

Institut Élie Cartan, Université Henri Poincaré-Nancy I, BP 239, 54506, Vandœuvre-lès-Nancy, France
*E-mail address*: Philippe.Chassaing@antares.iecn.u-nancy.fr
*URL*: http://www.iecn.u-nancy.fr/~chassain/

Institut Élie Cartan, Université Henri Poincaré-Nancy I, BP 239, 54506, Vandœuvre-lès-Nancy, France
*E-mail address*: Florent.Gillet@antares.iecn.u-nancy.fr

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden
*E-mail address*: svante.janson@math.uu.se
*URL*: http://www.math.uu.se/~svante/

Department of Computer Science, Illinois Institute of Technology, Stuart Building, 10 West 31st Street, Suite 236, Chicago, IL 60616-3729 U.S.A.
*E-mail address*: reingold@iit.edu

Institut Élie Cartan and LORIA, Université Henri Poincaré-Nancy I, BP 239, 54506, Vandœuvre-lès-Nancy, France
*E-mail address*: schott@loria.fr